Bovine and Rodent Tamm-Horsfall Protein (THP) Genes: Cloning, Structural Analysis, and Promoter Identification

HERON YU,¹ FEROZ PAPA,¹ AND VIKAS P. SUKHATME²

Renal Division, Department of Medicine, Beth Israel Hospital and Harvard Medical School, 330 Brookline Avenue, Boston, MA 02215

We have isolated bovine and rodent cDNA and genomic clones encoding the kidney-specific Tamm-Horsfall protein (THP). In both species the gene contains 11 exons, the first of which is noncoding. Exon/intron junctions were analyzed and all were shown to follow the AG/GT rule. A kidney-specific DNase I hypersensitive site was mapped onto a rodent genomic fragment for which the sequence is highly conserved in three species (rat, cow, and human) over a stretch of 350 base pairs. Primer extension and RNase protection analysis identified a transcription start site at the 3' end of this conserved region. A TATA box is located at 32 nucleotides upstream of the start site in the bovine gene and 34 nucleotides upstream in the rodent gene. An inverted CCAAT motif occurs at 65 and 66 nucleotides upstream of the start site in the bovine genes, respectively. Other highly conserved regions were noted in this 350 bp region and these are likely to be binding sites for transcription factors. A functional assay based on an in vitro transcription system confirmed that the conserved region is an RNA Pol II promoter. The in vitro system accurately initiated transcription from the in vivo start site and was highly sensitive to inhibition by α -amanitin at a concentration of 2.5 $\mu g/ml$. These studies set the stage for the further definition of *cis*-acting sequences and *trans*-factors regulating expression of the THP gene, a model for kidney-specific gene expression.

TAMM-HORSFALL protein (THP) is a glycoprotein with a molecular weight of 80 kDa, initially characterized by Tamm and Horsfall (1950) and subsequently purified from the urine of pregnant women by Muchmore and Decker (1985) as uromodulin. Pennica et al. (1987) showed that the two proteins were identical. THP is the most abundant protein in urine and is present in the kidneys of all placental mammals (Fletcher, 1972). Because one of our long-term research goals is to understand how gene expression is regulated in the kidney, we have chosen to use the THP gene as a model system for the following reasons. (1) The expression of THP is limited to the kidney in a nephron segment specific distribution. It is therefore a good example of a kidney-specific gene. Within the kidney, it is expressed only in the thick ascending limb of Henle's loop and

in the distal convoluted tubule (Schenk et al., 1971). THP is produced in all mammalian species with a similar nephron-specific localization. The gene is therefore likely to have conserved *cis* regulatory sequences, and comparison of homologous regions from different species could implicate motifs that might bind tissue-specific transcription factors. (2) THP mRNA is present in high amounts (0.5-1% of message). The abundance of this message should greatly facilitate studies of gene regulatory sequences.

As a first step toward elucidation of the mechanisms that regulate kidney-specific gene expression, we have isolated and characterized the bovine and rodent THP genes. These animal homologues of the human THP gene are desirable due to the availability of kidney tissue that is required for chromatin structural studies and for making nuclear extracts

Received April 4, 1994; revision accepted June 22, 1994.

¹Equal contribution.

²Address correspondence to Vikas P. Sukhatme.

(see below). In this article we describe the cloning of cDNAs containing the entire coding region and most of the noncoding region for bovine and rodent THP. The entire sequence of the cDNAs and the intron/exon structure of the two genes are also presented. We have identified a DNase I hypersensitive site in a highly conserved region that is also the promoter of this gene, as demonstrated by RNA mapping analysis and through an in vitro functional assay.

MATERIALS AND METHODS

Cloning, Restriction, and Sequence Analysis

cDNA libraries were in lambda gt10. The rodent library was a kind gift of Dr. G. Bell (University of Chicago). Standard procedures were used for cloning, restriction, and sequence analysis (Sambrook, 1989). Genomic libraries were in the lambda DASH II vector (Stratagene). cDNA and genomic inserts were cloned into pUC19 and pBluescript II SK, respectively.

RNA Isolation and Northern Analysis

RNA was isolated by the Chomczynski protocol (Chomczynski and Sacchi, 1987). Northern blotting and hybridization were as described (Current Protocols, 1991). Inserts from cDNA plasmids pBTHP1 and pRTHP1 were labeled by random priming (Feinberg and Vogelstein, 1983).

DNA Sequencing

The cDNA phage insert was subcloned in pUC19 and DNA sequencing was carried out using the chain-termination method with Sequenase (United States Biochemical).

DNase I Hypersensitive Site Analysis

Nuclei were isolated from 5 g of kidney and liver by mincing tissue in 25 ml of 0.25 M sucrose, 10 mM Tris HCl (pH 7.5), 5 mM NaCl, 3 mM MgCl₂, 0.5% NP-40. Tissue was homogenized in a Potter-Elvejem homogenizer and the nuclei were isolated by centrifugation. They were washed three times with the same buffer without NP-40 and resuspended at a DNA concentration of 1 mg/ml. DNase I was added to aliquots of the nuclei at 0, 1, 2, 5, and 10 μ g/ml final concentration and incubation was carried out for 2 min at 25°C. One volume of stop solution (12.5 mM EDTA, 0.5% SDS, 500 μ g/ml proteinase K) was added and incubation was carried out for 30 min at 37°C. Phenol-chloroform extractions were performed twice and a chloroform extraction was carried out before the DNA was precipitated with EtOH. The DNA was resuspended in TE and digested with Hind III, electrophoresed, and blotted onto nylon membranes. Prehybridization and hybridization was in 6 × SSC, 0.5% SDS, 100 μ g/ml salmon sperm DNA, 5 × Denhardt's reagent (Current Protocols, 1991), and 50% formamide. A random printed,³²P-labeled 300 bp intron 1 probe that was generated by PCR (diagrammed in Fig. 3) was used.

RNA Mapping Analysis

RNase protection assays were according to the method of M. Gilman (Gilman in Current Protocols, 1991). Total RNA (10 μ g) was used for all reactions. For bovine THP, an EcoRI-Pst I 800 bp genomic fragment that contains the first exon was subcloned into pBluescript II SK+. The plasmid was digested with Ssp I and the 220 bp fragment was transcribed by T₃ RNA polymerase to give an antisense riboprobe. For rat THP, a PCR fragment from the rat conserved region -350 to +35 was cloned into the SmaI site of pBluescript II SK. The plasmid was cut with Xhol and transcribed with T₃ RNA polymerase to give an antisense riboprobe; 5×10^5 cpm were used for the hybridization.

Primer extension analysis was carried out as described by R. Kingston (Current Protocols, 1991) with specific exon 2 oligos BTHP4 and BTHP7 for bovine samples and RTHPRE36 and RTH5 for rat samples. mRNA was prepared by using oligo-dT columns (Boehringer Mannheim). Fewer nonspecific extensions resulted when 2 μ g mRNA rather than 10 μ g total RNA was used.

In Vivo Transcription

Extracts for in vitro transcription were made by the method of Gorski (1986) with the following modifications for kidney extracts: homogenization buffer 1 was supplemented with 1% nonfat milk, 5 μ g/ml bestatin, 0.1 μ g/ml leupeptin, 0.1 μ g/ml pepstatin, 1 mM NaMO₄ (we found that inclusion of this phosphatase inhibitor was critical to the success of the procedure); homogenization buffer 2 contained the same supplement. Nuclear lysis buffer and nuclear dialysis buffer contained the same supplement but without the nonfat milk. Moreover, prior to homogenization, the kidneys were passed through a garlic press (Crate and Barrel worked best!).

The transcription reactions were essentially as



FIG. 1. (A) Northern blot of total cortical (C), medullary (M), and papillary (P) bovine RNA probed with the insert of pRTHP1. (B) Northern blot of total kidney (K) and liver (L) rat RNA probed with the same insert. Markers refer to 28S and 18S ribosomal sizes (\approx 5 kb and \approx 2 kb, respectively).

described (Gorski et al., 1986). Template DNAs were prepared by a triple-spin CsCl₂ procedure and used at final concentrations of 50 μ g/ml; 60 μ g/ of protein was used per assay and α -amanitin was at 2.5 μ g/ml.

RESULTS

Rodent and Bovine THP cDNA Homologues

A 550 bp BamH1-Hind III fragment from the human cDNA clone UM19 was used to screen rodent and bovine kidney cDNA libraries (Hession et al., 1987). Approximately 0.5% of plaques were positive in each library. The longest bovine clone, designated pBTHP1, had an insert of 2.5 kb whereas the longest rodent clone, pRTHP1, contained a 2.2 kb fragment. The inserts were used as probes for Northern analysis. As seen in Fig. 1A, the bovine probe hybridized strongly to a 2.6 kb transcript in outer medulla but weakly to cortex or papilla. The rodent probe hybridized to kidney but not liver (Fig. 1B).

pBTHP1 and pRTHP1 were fully sequenced

(Fig. 2). pBTHP1 contains 1929 bp in its longest open reading frame and pRTHP1 contains 1932 bp leading to a predicted molecular weight of THP of \approx 68 kDa (taking into account the 24 amino acidleader sequence cleaved from the precursor protein). Given the measured M.W. of \approx 90 kDa, the carbohydrate moiety accounts for about 25% of the total. At the nucleotide level the bovine sequence shows 72.5% and 80.0% homology to the rodent and human sequences, respectively. Pennica et al. (1987) noted an unusually high number of CpG dinucleotides in the human third exon. This is also seen in the bovine third exon (63 CpGs). However, only 29 CpGs were noted in the rodent third exon, still a number greater than the predicted number due to CpG suppression, which is 18. The significance of this CpG cluster is not known.

Genomic Clones Spanning the Entire Rodent and Bovine THP Genes

Rat genomic clones were isolated by screening approximately 2×10^6 recombinant phage from a rodent genomic library (Stratagene) using the pRTHP1 cDNA insert as a probe. This screening yielded 16

1 10 12 20 30 40 50 40 CCTTTCT86TTTCASGACTTCASGBATCASGBACABABBAAATCTCAABABGCCABCCT	1210 1220 1230 1240 1250 1240 TACCTECATEACABCCABTECTCABECTABABABGEGEACEBBACEBBATE TycleautisabspereBicysbergityPhetThefUldrg61yAspArgAsptrphetSer
70 80 90 100 110 120	1270 1280 1290 1300 1310 1320
	Val Val ThrProAl aArgAsp6l yProCys6l yThrVal Het ThrArgAsn6l uThrHis
130 140 150 140 170 180 AGTGTCTTTTCTCTCGAAGTCAATTCAATGGAATGGAAGGGAGGAGGAGCCTCTTGGGATCCATC LysCysLeuPheSerProAsnPheHetTrpHetAlaAlaValValThrSerTrpVallie	1330 1340 1350 1340 1370 1370 GCCACATACAGGACACACTICTACCTGGAGATGATCATCCGTGAGCTC AlaThrTyrSerAsnThrLeuTyrLeuAlaAspGlulleileileAgAspLeuAsnIle
190 200 210 (¹³ 220 230 240 ATACCTGCABCAACTGACAACTCATCGCTCTGCABTGCACGCAATGGCC 11@FroA1aAlaThrAspThrSerSerAlaLysSerCysSerGluCysHisSerAsnAla	1390 1400 1410 1420 1430 1440 Asaatcaactittessisetectetastasaastaskasteasectsaaaste ArgilaasnPheAlacysSertyrProLeuAspHetLysValSerLeuLysThrSerLeu
250 260 270 280 290 300 ACTTBTACBGTGGACBGGGCTGCCACGAGCTGCCAGGAGGGCTTCACTGGCGAC ThrCysThrVal AspGl yAl aAl aThrThrCysAl aCysGl nGl uGl yPheThrGl yAsp	1450 ↓ 1460 1470 1480 1490 1500 CAGCCAATGGTCAGTGGCCCTCAACATCAGCATGGGCGGGACCGGCACATTCACCGTGCGA GInProMetValSerAlaLeuAsnileSerMetGlyGlyThrGlyThrPheThrValArg
310 320 330 340 330 360 GGCCTCSAG15TG16GATC19GACGAATGCSCCGTTC16G6G6GCGCACAAC16CTCCGCC G1 yLeuG1uCysVa1AspLeuAsp61 uCysA1 aVa1LeuG1 yA1 aHi sAsnCysSerA1 a	1510 1520 1530 1540 1550 1560 ATGGCACTCTTCCAGAGCCCTGCCTACCACAGCCCTACCAAGGCTCCTCTGTGACCCTG Met Al a Leuphegin Ser ProAl a Tyr ThrGin ProTyr Gingiy Ser Ser Val Thr Leu
370 380 390 400 410 420 ACCAAGAGCTGCGTGAATACGCTGGGCTCTTACACGTGGCTCTGCCCTGAAGGTTTCTC ThrLysSerCysValAsnThrLeuGlySerTyrThrCysValCysProGluGlyPheLeu	1570 1580 1590 1600 1610 1620 TCCACAGAAGEGTTTCTCTACGTCGGCACCATGCTGGAGGGGGGGGCTTGTCCCGGTTT SerThrG1uAlaPheLeuTyrVa131yThrMetLeuAspG1yG1yAspLeuSerArgPhe
430 440 450 460 470 480 CTGAGCTCGBAGCTGGGCTGCGAGGATGTGGAGAGCCAGGGCTCAGCCGC LsuSerSerGiuLeuGiyCysGiuAspValAspGiuCysAlaGiuProGiyLeuSerArg	1630 1640 1650 1660 1670 1680 GTACTGCTCATGACCAACTGCTATGCCACACCCAGCAGCAATGCCACAGACCCCTTGAAA ValLeuLeuMetThrAsnCysTyrAlaThrProSerSerAsnAlaThrAspProLeuLys
490 500 510 520 530 540 TBCCACGCCCTGBCCACTTGCATCAATBBCGAGGGCAACTACTCGTGCGTGTGTCCTBCG CysHisAlaLeuAlaThrCysIleAsnBlyGluGlyAsnTyrSerCysValCysProAla	1690 1700 8 1710 1720 1730 1740 TACTTCATCATCCAGGACAGATGTCCACGTGCTGCGGACTCAACCATCCAAGTGGAGGAG TyrPheIleIleGinAspArgCysProArgAlaAlaAspSerThrlleGinValGluGlu
550 560 570 580 570 600 GGCTACCTG3GAGACBGAAGGCACTGTGAGTGTTCCCCGGGCTCCTGTGGGCCTGGGCTA Gi yTyrLeuGi yAspGi yArgHi sCysGi uCysSerProGi ySerCysGi yProGi yLeu	1750 1760 1770 1780 1790 1800 AATGGGGAGTCCCCTCAGGGCCGGTTTTCTGTCCAGATGTTCCGTTTTGCTGGAAACTAC AsnG1yG1uSerProG1nG1yArgPheSerValG1nHetPheArgPheAlaG1yAsnTyr
610 620 630 640 650 660 GACTGCGTGCGGGGGGGGGGCGCCGCGGGGGCCCCGTGCCAGGTGCACCGCATC AspCysValArgGluGiyAspAlaLeuValCysValAspProCysGlnValHisArgIle	1810 1820 1830 1840 1850 1840 GACCTGBTGTACCTGCATTGTGAAGTGTATCTCTGTGACACCGTGAATGAA
670 680 690 700 710 720 CTGGACGAATACTGGCGCAGCACAGAGTACGGCTCCGGCTACATCTGTGATGTCAGTCTG LeuAspGluTyrTrpArgSerThrGluTyrGlySerGlyTyrIleCysAspValSerLeu	9 1870 1880 1870 1900 1910 1920 CCTACCTGCCCTGAGACCAGATTCCGCAGTGGGAGCATCATAGACCAAACCCGTGTCCTG ProThrCysProGluThrArgPheArgSarGlySerIlelleAspGlnThrArgValLeu
730 740 750 760 770 780 GGCGGCTGGTACCGCTTCGTGGGCCAGGCCGGCGGCGGCCCGCCTGCCT	1930 1940 V 1950 1960 1970 1960 AACTTEGGTCCCATCACAGGAAGGGGGGCCAGGCTGCAATGTCAAGGGCTGCTCCCAGT AsnLeuGiyProlleThrArgLysGiyGiyGinAiaAiaMetSerArgAiaAiaProSer 11
790 800 810 820 830 940 GTCCTGCACCGCCGCGCGCGCGCCCTATGTGGCTCACCGGCCGCATCCATC	1990 2000 2010 2020 2030 2040 AGCTTGGGGGCTTGCAGGTCTGGCTGCCTCTGCTGGGCCACTTTGACCCTGATG SerLeuGlyLeuLeuGlnValTroleuProLeuLeuLeuSerAlaThrLeuThrLeuMet
850 860 870 880 890 900 GAGGGCATCGTGAACCGCGTGGCCTGTGCACACTGGAGCGGTGACTGCTGCGGGC GlußiyileVelAsnArgVelAlaCysAlaHisTrpSerGlyAspCysCysLeuTrpAsp	2050 2060 2070 2080 2090 2100 TCTCCGTGACTGTGGCCGGAAATCCTGTACTCTGTGGCTACCAAACTCACTTTCTACTGA SerPro
910 920 930 940 950 960 GCGCCCATCCAAGTGAAGGCCTGTGCGGGGGGCTACTACGTGTACAACCTGACGGCGCCCC Al aProlleGinValLysAl aCysAl aGiyGiyTyrTyrValTyrAsnLeuThrAlaPro	2110 2120 2130 2140 2150 2160 CTGCGAGGCCATGCAGGCTAGCACCCCATAGAGAAAGGAGCCCACACTTCAGTCAC
970 980 970 1000 1010 1020 CCTGAGTGCCATCTGGCTTACTGCACAGACCCCCAGCTCTGTGGAGGGGACGTGTGAGGAG ProGluCysHisLeuAlaTyrCysThrAspProSerSerValGluGlyThrCysGluGlu	2170 2180 2190 2200 2210 2220 TCTBCTCCTATTTTTCTCCCCTTTAATCCTTGTTACCAAAAAAATABCCTGTGTCTTTAA
1030 1040 1050 1060 1070 1080 TGCC6TGTGGATGAGGACTGCAAATCGGATGGCACTGCCAGTGCAAACAG CysArgValAspGluAspCysLysSerAspAsnGlyGluTrpHisCysGlnCysLysSin	2230 2240 2250 2260 2270 2280 ATSCTSCTTTCTCTCAAAATBGGTCTTGTGATATGTCTGTTCATGAGGCCCCTATCTCCT
1090 U 100 1110 1120 1130 1140 GACTTCAACGTCACCGATCTCTCCCCCCCGGAGGCCGGAATGTGGGGTTGATGAC AspPheAsnValThrAspLeuSerLeuLeuGluArgArgLeuGluCysGlyValAspAsp	2290 2300 2310 2320 2330 2340 Талабевтаттесалаатаатаатталалаластсавтсталалалалалалалала
1150 1160 1170 1180 1190 1200 ATTAAGTTGTCCCTGAGCAAGTGCCAGCTGAAGAGTCTGGGAAGGGTCTTCATG IlelysLeußerLeußerLysCysGlnLeulysSerLeußtyPheGluLysValPheHet	

FIG. 2. (A) Complete sequence of the insert of pBTHP1. Exon boundaries have been denoted.

positive clones. The clones were further characterized and subgrouped by plaque hybridization with 17-mer oligos corresponding to selected exon sequences. Three clones were found to overlap and contained all 11 exons. Restriction blot analysis of these clones showed that the gene spans approximately 16 kb (Fig. 3B). The pBTHP cDNA insert was similarly used to screen 1×10^6 phage of a bovine genomic library (Stratagene). Out of eight positive clones, three were chosen for further structural analysis by restriction blot hybridization with four fragments of pBTHP1 (Fig. 3A).

The bovine gene spans 25 kb. Exon/intron junc-

tions were assigned to both genes by sequence analysis using the junctions reported by Pennica et al. as a guide. All 10 exon/intron junction sequences follow the AG/GT rule (Breatnach and Chambon, 1981). The junctions show complete homology to the human sequences. Both the rat and bovine THP genes consist of 11 exons, the first of which is noncoding.

A Kidney-Specific DNase I Hypersensitive Site at the THP Gene Promoter

Because promoters are frequently associated with DNase I hypersensitive sites (Gross and Garrard, к

. 10³

120 TCT

	.6					
	1310	1220	1230	1240	1250	1260
	µ210	1220	1230	1240	1200	CATC
TTGAGG	AGAAATGAAA	NCCCACGCCAC	CTACAGCAA	CACCETETACU	I GOLAAG I GA	BHIC
LeuAro	ArgAsnGlul	[hrHisAlaTh	rTyrSerAs	nThrLeuTyrl	.euAlaSerGl	ulle
	•		•			
	1370	1280	1280	1300	1310	1320
	12/0	1200	1270	1000	CTOTOCOCOT	0000
ATCATC	AGGGATATCA	ACATCAGAA	CAACTIIGA	ATGUTUTIALU	LICIGGHCHI	GHHH
llelle	ArgAspileA	AsnIleArgIl	eAsnPheGl	uCysSerTyrf	roleuAsphe	itLys
				7		
	1330	1340	1350	1360	1370	1380
CTC+CT			TATCOTTAC	,	TCAGCTTEEG	TORO
GICHGI	CIGAHGHUU	LUCIALAGU	THIOSTING	I GCC I I GMMCF		
ValSer	LeuLysThr§	SerLeuGlnPr	•oMetValSe	rAlaLeuAsni	[]eSerLeuGl	YGIY
	1390	1400	1410	1420	1430	1440
ACA660	AAGTTCACTO	TREARATER	ACTGTTCCA	GAACCCCTACC1		CTAC
Incory	Cyseneines	ALOINNETH	erentuent	negneroinei	i yr ffir Grifer	0.91
	1450	1460	1470	1480	1490	1500
CAAGGT	CCTTCTGT64	ATECTOTICAC	CTGAGGCTTT	TCTGTATGTG	GCACCATGC1	GGAT
61 661	ProSecuals	tetl au Car Th	-G141	al au Tya Val (1 TheHetl	
G11101)	ri user vali	Herceuser II	il or unit arm			
	1510	1520	1530	1540	1550	1560
666661	GACTTGTCC	CGGTTTGTAC	TECTAATGAC	CAACTGCTAT	SCCACACCCAG	STAGC
61 461 4	Asol auSer	AroPhevall	iul auMat Th	-AsnCvsTvr/	al a Thr ProSe	r Ser
•• , •• ,						
				18		
	1570	1560	1390	1600 🖵 🗸	1619	1620
AACTCO	ACGGACCCT	STAAAATACT'	TCATTATCCA	GGACAGATGT	CACATACAGA	AGAT
AsoSer	ThresoPro	Jall veTvrPl	nellelleßi	AsparoCysl	ProHisThrGI	UASO
	The report of			Internet Acha		
	1630	1640	1650	1660	1670	1960
ACAACO	ATTCAGGTG/	ACAGAGAATG	SCGAGTCCTC	TCAGGCCCGG'	TCTCTATTC	AGATG
The The	TieGioVal'	hrG1.AsnG	vGluSerSe	rGinAlaAroi	PheSer IleGi	oMet
				•		
	1690	1700	1/10	1720	1/30	1740
TTCCBB	TTTCCACCA	*********				
		MMLILLBALL	INGICIACU	TCACTGCGAG	STGIACCIGIC	31044
PheArc	PheAlaGlv	Aso Ser AsoLi	NUVALTVELO	uHisCvsGlu	STGIACCTGTG Val TvrLeuCy	51 GAC
PheArg	PheAlaGly	AshSerAspLi	euValTyrLe	TCACTGCGAGO uHisCysGlu	Val TyrLeuCy	si GAC
PheArg	PheAlaGly	AsnSerAspLi	euValTyrLe	uHisCysGlu	Val TyrLeuCy	shap
PheArg	PheAlaGly	1760	1770	TCACTGCGAGG uHisCysGlu 1780	Al TyrLeuCy 1790	1800
PheArg	PheAlaGly 1750 BAGTGAGCAG	1760	1770	TCACTGCGAGG uHisCysGlu ¹ 1780 TACTAGATAT	1790 CGAAGTGGGAA	1800 CTTC
ACTATE	PheAlaGly 1750 BAGTGAGCAG SerGluGln	1760 TGTAAACCTA	1770 CCTGTTCTGG	TCACTGCGAG(uHisCysGlu' 1780 TACTAGATAT(yThrArgTyr)	1790 GAAGTGGGAA Ya Ser Gi yaq	1800 ACTTC
ACTATE	PheAlaGly 1750 SAGTGAGCAG SerGluGln(1760 TGTAAACCTAI CysLysProTi	1770 CCTGTTCTGG DrCysSerG1	TCACTGCGAG(uHi sCysG1u 1780 TACTAGATAT(yThrArgTyr)	1790 CGAAGTGGGAA ArgSerGlyAa	1800 ACTTC
ACTATE	PheAlaGly 1750 BAGTGAGCAG SerGluGln(1760 TGTAAACCTAI CysLysProTi	1770 CCTGTTCTGG	TCACTGCGAG(uHisCysGlu' 1780 TACTAGATAT(yThrArgTyr)	1790 CGAAGTGGGAA ArgSerG1yAt 110	1800 ACTTO
ACTATE ThrMet	1750 AGTGAGCAG SerGluGlni 1810	1760 TGTAAACCTA CysLysProTi	1770 CCTGTTCTGG TCYsSerG1 1830	TCACTGCGAG uHisCysGlu' 1780 TACTAGATATI yThrArgTyri 1840	1790 CGAAGTGGGAA ArgSerG1yAn 18850	1800 ACTTC 1860
ACTATO ACTATO	1750 3AGTGAGCAG SerGluGini 1810	1760 TGTAAACCTA CysLysProTi 1820	1770 CCTGTTCTGG nrCysSerG1 1830	TCACTGCGAGG uHisCysGlu' 1780 TACTAGATAT(yThrArgTyr/ 1840 CACACGACGACAA	1790 1790 CGAAGTGGGAA ArgSerGlyAt 1 10 19850 GGGTCCAGG	1800 1800 ACTTC 1860
ACTATE ACTATE ThrMet	1750 3AGTGAGCAG SerGluGln(1810 ICAGACCCGT)	ARCTOLIGACO	1830	TCACTGCGAGG uHisCysG1u 1780 TACTAGATATI yThrArgTyri 1840 CACACGACGACAA	1790 CGAAGTGGGAA ArgSerG1yAt 10 1850 GGTGTCCAGGG	1800 1800 ACTTC 1860 1860
ACTATO ThrMet ATAGAT	1750 3AGTGAGCAG SAGTGAGCAG SarGluGini 1810 FCAGACCCGTI SGInThrArg	ABDSERASDLA 1760 TGTAAACCTA CysLysProTI 1820 GTCCTGAACT ValLeuABDL	1830 1830 1830 1930 1930 1930 1930 1930 1930 1930 19	TCACTGCGAGG uHisCysGluv 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAG eThrArgGlui	1790 CGAAGTGGGAA ArgSerGlyAn 100 5050 GGTGTCCAGG 51yValGlnA	1800 ACTTO IB60 IB60 CCTCA
ACTATE ThrMet ATAGA1 IleAsp	1750 3AGTGAGCAG SerGluGln(1810 ICAGACCCGT) SGInThrArg	ARCTELEREL ABANSERABLI 1760 TGTAAACCTAI CYSLYSPROTI 1820 GTCCTGAACT ValleuABALI	1000 1770 1770 CCTGTTCTGG nrCysSerG1 1830 TGGGTCCCAT euGlyProI1 11	TCACTGCGAGG uHisCysGlu 1780 TACTAGATATG yThrArgTyri 1840 CACACGACAAH eThrArgGlni	1790 CGAAGTGGGGAA ArgSerGlyAr 1850 GGTGTCCAGGG SlyValGlnA	1800 ACTTO ISOPhe 1860 CCTCA
ACTATE ThrMet ATAGAT IleAsp	1750 3AGTGAGCAG SerGluGin(1810 FCAGACCCGT(5GInThrArg) 1870	ABDSFRADL	1900 ALTYRLE 1770 CCTGTTCTGG TCYSSerG1 1830 TGGGTCCCAT BUGI yProI1 1970	TCACTGCGAGG uHisCysGluv 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAG eThrArgGlni 1900	1790 CGAAGTGGGAA ArgSerGlyAa 18850 GGTGTCCAGGG GIyValGlnA 1910	1800 1800 ACTTC INPhe 1860 CCTCA LaSer 1920
ACTATO ThrMet ATAGAT IleAsp GTGTCC	1750 1750 3AGTGAGCAG SerGluGlni 1810 ICAGACCCGTi GlnThrArgi 1870 2AAGGCTGCT	ABD STREAM 1750 1750 TGTAAACCTAI CYSLYSProTI 1820 GTCCTGAACT ValLeuASDL 1880 TCCAGCAACT	1010 CALCH 1770 CCTGTTCTGG 1030 10 1030 1000 1000 1000 10	TCACTGCGAGG uHisCysGluv TACTAGATATI yThrArgTyri 1840 CACACGACAAI eThrArgGluv 1900 GAGCATCTGG	1790 CGAAGTGGGA/ ArgSerGlyAn 100 1850 GGTGTCCAGGG SlyValGlAA 1910 CTGCTGCTGCTGT	1800 ACTTC IB60 CTTC IB60 CCTCA IB60 CCTCA IB60 CCTCA IB60 CCTCA IB60 CCTCA
ACTATO ThrMet ATAGAT IleAsp GTGTCC	1750 BAGTGAGCAG SerGluGlad 1810 CCAGACCCGTI JGIaThrArg 1870 CAGGCTGCT	ASING CLOBACC ASING ASING ASIN	1010 CALCH 1770 CCTGTTCTGG 1030 TGGGTCCCAT auGlyProll 11 10 10 10 10 10 10 10 10 10	TCACTGCGAGG uHisCysGlu' 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAG eThrArgGlni 1900 'GAGCATCTGG uScr11gTcn	Val TyrLeuCy 1790 CGAAGTGGGAA rgSerGlyAt 100 8850 GGTGTCCAGGG SlyValGlnA: 1910 CTGCTGCTGTT	1800 ACTTC INPhe 1860 CCTCA LaSer 1920
ACTATO ACTATO ThrMet ATAGA IleAs GTGTCO ValSer	1750 SaGTGAGCAG SerGluGloi 1810 FCAGACCCGT GloThrArg 1870 CAGGCTGCT LysAlaAla	AAAN SERASPLA AAAN SERASPLA 1740 (STADACCTAI TGTAAACCTAI CysLysFroll 1820 GTCCTGAACT ValleuAshL 1880 TCCAGCAACT SerSerAshL	UVal TyrLe 1770 CCTGTTCTGG TrCysSerG1 1630 TG6GTCCCAT 40G1 yPro11 111 1670 1660 GTTCCT auG1 yPheLe	TCACTGCGAGG uHi sCysGlu' 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAI GACATCGGA 1900 GAGCATCTGG uSer IleTrpi	Val TyrLeuCy 1790 CGAAGTGGGAA ArgSerG19At 10 1950 GGTGTCCAGG(SI yVal G1nA) 1910 CTGCTGCTGTT LeuLeuLeuPi	1800 ACTTC IB60 CTTC IB60 CCTCA IASer 1920 TTCTC NeLeu
ACTATE ACTATE ThrMet ATAGA1 IleAsp GTGTCC ValSer	1750 SAGTGAGCAG SerGluGlad (Blo (CAGACCCGT) SGlaThrArg (1870 SAGGCTGCT (LysAlaAla)	AAS SERASPL AAS SERASPL 1740 \$ 1760 \$ 1760 \$ 1760 \$ 1820 GTCCTGAACT ValLeuAsnL 1880 TCCAGCAACT SerSerAsnL	UValTyrLe 1770 CCTGTTCTGG nrCysSerGl 1830 TGGGTCCCAT euGlyProll 111 1970 TGGGGTTCCT euGlyPheLe	TCACTGCGAGG uHisCysGlut 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAA eThrArgGlut 1900 GAGCATCTGG GAGCATCTGG	Val TyrLeuCy 1790 CGAAGTGGGAA ArgSerGlyAn 100 8850 66TGTCCAGGG 51 yVal GlnAi 1910 CTGCTGCTGT 1910	1800 ACTTC INPhe 1860 CCTCA LaSer 1920 TTCTC
ACTATE ACTATE Thr Met ATAGA1 II eAsp GTGTCC Val Ser	PheAlaGly/ 1750 SaGTGAGCAG SerGluGlni (1810 CCAGACCCGT) GlnThrArgi 1870 CAAGGCTGCT -LysAlaAla	ARTECLARCE 1760 1760 TGTAAACCTAI CysLysProTi 1820 GTCCTGAACT ValLeuAsnL 1880 TCCAGCAACT SerSerAsnL 1740	UVal TyrLe 1770 CCTGTTCTGG nrCysSerG1 1830 1630 1660 1670 1770	TCACTGCGAGG uHi sCysGlu' 1780 TACTAGATATI yThr ArgTyri 1840 CACAGGACAAA eThr ArgGlni 1900 GAGCATCTGG uSer Il eTrpi 1960	Val TyrLeuC) 1790 CGAAGTGGGAA ArgSarG19At 100 1850 GGTGTCCAGG GGTGTCCAGG GGTGTCCAGG 1910 1910 1910 1910 1970	1800 CTTC INPhe 1860 CTTCA LaSer 1920 TTCTC NeLeu 1980
ACTATE ACTATE ThrMet ATAGA1 11 eAs GTGTCC Val Ser TCABC	1750 1750 1750 1860 1810 ICAGACCCGTI GlaThrArg' 1870 24060CTGCT -LysAlaAla 1930 1930	ARSERASLA AsaSerAspLi 1740 1 1740 1 1740 1 1740 1 1820 GTCCTGAACT ValLeuAsnLi 1880 TCCAGCAACT SerSerAsnL 1740 CTGATGGTTC CTGATGGTTC	INTELACT UValTyrLe 1770 CCTGTTCTGG nrCysSerGl 1630 TGGGTCCCAT UGIYProll 111 1670 TGGGGTTCCT auGiyPheLe 1950 ATTGATGGAP	TCACTGGGAGG uHisCysGlut 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAA eThrArgGlnt 1900 GAGCATCTGG uSer IleTrpi 1960 1960	Val TyrLeuC) 1790 CGAAGTGGGAA ArgSerGlyAn 100 1650 GGTGTCCAGG GGTGTCCAGG SI yValGlAA 1910 CTGCTGCTGCTGT 1970 CTGGTGTGGCC	1800 1800 CTTC IB60 CTCA 1860 CTCA 1860 CTCA 1920 TCCCA 1980 TCCCA
ACTATE ACTATE Thr Met ATAGA1 II eAsp GTGTCC Val Ser TCAGCC	1750 BAGTGAGCAG Ser GluGini 1810 rCAGACCCGTi JGIn Thr Ar gi 1870 SAAGGCTGCT LysAlaAla 1930 rACTTTGACC	ARTECLARCE ARTSerAspLi 1760 TGTAAACCTAI CysLysProTi 1820 GTCCTGAACT ValLeuAsnLi 1880 TCCAGCAACT SerSerAsnLi 1740 CTGATGGTTC LeuMetValH	INTELACT UVALTYLe 1770 CCTGTTCTGG DrCysSerG1 1830 1630 1660GTCCCAT 101 1670 1660GTCCCAT 101 107 1050 ATTGATGGAG	TCACTGCGAGG uHisCysGlu ¹ 1780 TACTAGATATI yThrArgTyri 1840 CACAGGACAAA eThrArgGlni 1900 GAGCATCTGG uSerIleTrpi 1960 MACAGAAAAAC	Val TyrLeu(C) 1790 CGAAGTGGGAA ArgSarG19At 100 18850 GGTGTCCAGG GGTGTCCAGG GGTGTCCAGG 1910 1910 1760 CTGCTGCTGTT LeuLeuLeuPi 1970 CTGGTGTGGGC	1800 ACTTC INPhe 1860 CCTCA 1860 CCTCA 1920 TCCCA
ACTATE ACTATE ThrMet ATAGA1 11 eAs GTGTCC Val Ser TCAGC Ser Al 4	1750 SAGTGAGCAG SerGluGini 1810 IGGATCCGT GlaThrArg 1870 CAGGCCGCT -LySAlaAla 1930 1930 ractTGACC aThrLeuThri	ARTECUSALC AnserAspL 1760 16TARACCTA CysLysFroTI 1920 GTCCTGAACT ValLeuAsnL 1980 TCCAGCAACT SerSerAsnL 1940 CTGATGGTTC LeuMetValH	UVALTYLA 1770 CCTGTTCTGG rCVsSerG1 1830 TGGGTCCCAT uG1YPro11 111 1970 TGGGGTTCCT auG1YPheLa 1950 ATTGATGGAF 15	TCACTGGGAGG UHISCYSGIU 1780 TACTAGATATI YTN ArgTyri 1840 CACAGGACAAI GAGCATCTGG GAGCATCTGG USGCIIeTrpi 1960 JAACAGAAAAAC	Val TyrLeuC) 1790 CGAAGTGGGAA rgSarG1yAt 10 9850 GGTGTCCAGG(1910 CTGCTGCTGT 1910 CTGCTGCTGT 1970 CTGGTGTGGCC	1800 ACTTC IB60 CCTCA IB60 CCTCA IS60 CCTCA IS60 TCCCA
ACTATE ThrMet ATAGAT IleAsp GTGTCC ValSer TCAGC SerAl	1750 BAGTGAGCAG Ser GluGlad 1810 rCAGACCCGTI JGIa ThrArgi 1870 SAAGGCTGCT LysAlaAlai 1930 rACTTGACC aThrLeuThri	ARTECUMPLE AnserAspL 1760 (TGTAAACCTA CysLysFroTI 1820 GTCCTGAACT ValLeuAsnL 1880 TCCAGCAACT SerSerAsnL 1940 CTGATGGTTC LeuMetValH	INTELACT UVALTYLe 1770 CCTGTTCTGG DrCysSerG1 1630 TGGGTCCCAT UG1YP 1670 TGGGTTCCT UG1YPALe 1950 ATTGATGGAG	TCACTGCGAGG UH1 SCysG1u ¹ 1780 TACTAGATATI yThr ArgTyri 1840 CACACGACACAAA eThr ArgG1ni 1900 GAGCATCTGG USer II eTrpi 1960 MACAGAAAAAC	Val TyrLeuc) 1790 CGAAGTGGGAT TGSerG1yAi 10 1950 GGTGTCCAGG SI YVal G1nA: 1910 CTGCTGCTGT 1970 CTGGTGTGGC 2020	1800 ACTTC INPhe 1860 CCTCA ISSer 1920 TTCTC INFLEU 1980 TCCCA
Phear c ACTATO Thr Met ATAGA1 Il eAsc GTGTCC Val Ser TCAGC Ser Al J	1750 Segre Lagrid Segre Lagrid	And Techanic respiration of the second secon	UVal TyrLe 1770 CCTGTTCTGG rCysSerG1 1830 TGGGCCCCAT uG1 yPro11 111 1970 TGGGGTCCCAT GGGGTCCCAT GGGGTCCCAT 1950 ATTGATGGAA 15 2010	TCACTGCGAGG UH1 SCys61u ¹ 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAAA eThrArg61ni 1900 GAGCATCTGG uSer I1 eTrpi 1960 MACAGAAAAAC 2020	Val TyrLeu() 1790 CGAAGTGGGAA Ar gSerGl YAH 190 B 850 GGTGTCCAGG(1910 CTGCTGCCAGG(1910 CTGCTGCTGT LeuLeuLeuPH 1970 CTGGTGTGGCC 2030	1800 ACTTC INPhe 1860 CCTCA 1860 CCTCA 1920 TCCCA 1920 TCCCA 2040
ACTATO ACTATO ThrMet ATAGAT II eAsp GTGTCC Val Ser Ser Al . GTTCAG	1750 1750 1967 19750 1967 1970 107 107 107 107 107 107 107 1	ARTECUSACC AnserAspin 1760 (TGTARACCTAI CysLysFroll 1820 GTCCTGAACT ValLeuAsnL 1880 TCCAGCAACT SerSerAsnL 1940 CTGATGGTTC LeuMetValH 2000 GCCAGGAGTG GCCAGGAGTG	UVal TyrLe 1770 CCTGTTCTGG TGGGTCCCAT UG1 YProll 1630 TGGGTCCCAT UG1 YProll 1670 UG1 YProll 1670 1770	TCACTGGGAGG UHI SCYSGIU' 1780 TACTAGATATI YTH ArgTyri 1840 CACACGACAAAA eThrArgGIni 1900 GAGCATCTGG USSerIIeTrpi 1960 MACAGAAAAAC 2020 CTGGTAAGCA	Val TyrLeuC) 1790 CGAAGTGGGAY TGSerG1YA1 10 1950 GGTGTCCAGG GGTGTCCAGG GGTGTCCAGG 1910 CTGCTGCTGCT 1970 CTGGTGTGGGC 2030 GCCAGGAGAGG	1900 1900 1900 1900 1900 1920 1920 1920 1920 1920 1920 2040 AGGG4
ACTATO ThrMed ATAGAI IleAsp GTGTCC ValSer TCABC SerAld GTTCA	1750 1750 1750 186TGAGCAG 1810 1810 1870 1870 1870 1870 1870 1970 1970 1970 1970 1970 1970 1970 1970 1970 1970	ARTECOMPLE AnserAspli 1740 [S TGTARACCTAI CysLysFroTi 1820 GTCCTGAACT ValLeuAsnL 1880 TCCAGCAACT SerSerAsnL 1940 CTGATGGTTC LeuMetValH 2000 GCCAGGAGTG	UVal TyrLe 1770 CCTGTTCTGG TCYSSErG1 1830 TGGGTCCCAT UG1YPr011 1970 TGGGTTCCT UG1YPr011 1970 TGGGTTCCT UG1YPhLe 1950 ATTGATGGAG 2010 GGGATGCAGG	TCACTGGGAGA UHI SCYSGIU' 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAA eThrArgGIni 1900 GAGCATCTGG 1940 IAACCAGAAAAAC 2020 CCTGGTAAGCA	Val TyrLeuC) 1790 CGAAGTGGGAT GGAGTGGGAT 1910 1910 CTGCTGCTGCT 1910 CTGCTGCTGCTGCT 2030 GCCAGGAGAGAG	1800 ACTTC INPhe 1860 CCTCA INPhe 1860 CCTCA INPhe IN
ACTATO ThrMet ATAGAT II eAsp GTGTCC Val Ser TCAGC Ser Al	1750 BAGTGAGCAG SerGluGland 1810 FCAGACCCGT DGInThrArg 1870 CAGAGCCGCT 1930 FACTTTGACC ThrLeuThr 1930 FACTTTGACC ThrLeuThr 1970 CTTCCTGCTG	ARTECUSALC AnserAspli 1760 (TGTARACCTAI CysLysFroll 1820 GTCCTGAACT ValLeuAsnL 1880 TCCAGCAACT SerSerAsnL SerSerAsnL 1940 CTGATGGTTC LeuMetValH 2000 GCCAGGAGTS	UNATYTLE 1770 CCTGTTCTGG TCYSErGI 1830 TGGGTTCCCAT SUGIYPHLE 1950 ATTGATGBAP 1950 ATTGATGBAGGAGG 2010	TCACTGGGAGG UHI SCYSGIU' 1780 TACTAGATATI YTH ArgTyr/ 1840 CACACGACAAA eThrArgGIni 1900 GAGCATCTGG GAGCATCTGG 1960 MACAGAAAAAC 2020 CCTGGTAAGCA	Val TyrLeuC) 1790 CGAAGTGGGAN 190 1950 GGTGTCCAGG GGTGTCCAGG 1910 CTGCTGCTGT 1910 CTGCTGCTGT 1970 CTGGTGTGGGC 2030 GCCAGGAGAGG	1800 ACTTC ISO ACTTC ISO CCTCA ISS ISS ISS ISS ISS ISS ISS ISS ISS IS
ACTATO ThrMet ATAGA1 IleAsp GTGTCC ValSer TCAGC SerAld GTTCAC	1750 1750 1967 1970 1910 1910 1970 1070	ARTECOSACT AnserAspin 1760 [S TGTAGAACCTAI CysLysFroTi 1820 STCCTGAACT ValLeuAsnLi 1880 TCCAGCAACT SerSerAsnL 1940 CTGATGGTTC LeuMetValH 2000 GCCAGGAGTG GCCAGGAGTG	uveltyrLe 1770 CCTGTTCTGG 1830 TGGGTCCCAT suGiyProll 1970 1950 ATTGATGGAF 2010 GGGATGCAGC 2070	TCACTGGGAGG UH1 SCYSG1u' 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	2030 2030 2030 2030 2030 2030 2030 2030	1800 ACTTC INPhe 1860 CCTCA 1860 CCTCA 1860 CCTCA 1920 TCCCCA 2040 AGGG4
Phearc ACTATE ThrMet ATAGAT 11eAs GTGTCC ValSer TCAGC SerAl GTTCAG	1750 BAGTGAGCAG SerGluGial (B10) (CAGACCCGTI GlaThrArg) 1870 CAGGCCGCT -LysAlaAla 1930 rACTTGACC AThrLeuthri 1990 CTTCCTGCTG 2050	ARTECLARCE ARTSERARCTAR 1740 [S TGTARACCTAR CysLysFroTI 1820 GTCCTGAACT VALLEUARNL: 1880 GTCCAGCAACT SerSerARNL 1940 CTGATGGTTC LeuMetValH 2000 GCCAGGAGTG 2040	UNITYLE 1770 CCTGTTCTGG TCYSErGI 1830 TGGGTCCCAT SUGIYProll 1970 TGGGTTCCTGGAG 1950 ATTGATGGAG 2010 2070 2070	TCACTGCGAGG UHI SCYSGIU' 1780 TACTAGATATI YThr ArgTyr/ 1840 CACACGACAAAAA eThr ArgGIni 1900 GAGCACTCGG GAGCACTCGG USer II eTrpi 1960 MACAGAAAAAC 2020 SCTGGTAAGCA 2080	Val TyrLeuC) 1790 CGAAGTGGGAA rgSerGlyAt 10 10 10 10 10 10 10 10 10 10	1900 ACTTC SnPhe 1800 ACTTC SnPhe 1860 CCTCA 1860 CCTCA 1920 TCCCA 1920 CCTCA 1920 CCCCA 1920 CCCCA 2040 AGG64
ACTATO ThrMet ATAGAT IleAsp GTGTCC ValSer TCAGC SerAld GTTCAG	1750 1750 1967 1975 1975 1970 1070	ARTECLARCE 1760 [1760 [16760 ASPC 16760 ASPC 1820 5770 ASPC 1820 5770 ASPC 1880 1000 5780 ASPC 1940 CTGATGGTTC Lew Hetval H 2000 GCCAGGAGTB 2060 AGCTTGCACT	2010 20170 21770 21770 201770 201770 2018 2019 2010 2010 2010 2010 2010 2010 2010	TCACTGGGAGG UH1 SCYSG1u ¹ 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAA eThrArgG1ni 1900 GAGCATCTGG USer II eTrpi 1940 MACAGAAAAAC 2020 SCTGGTAAGCA 2080 CTTTAACCCT	2030 2030 2030 2030 2030 2030 2030 2030	1800 ACTTC SnPhe 1860 CCTCA 1860 CCTCA 1920 TTCTC neLet 1980 TCCCA 2040 ACG66 ACC66
ACTATE ACTATE ThrMet ATAGAT 11 eAs GTGTCC Val Ser TCAGC SerAl GTTCA	1750 300TGAGCAG 300TGAGCAG 300TGAGCAG 1810 100 1810 100 1870 300 300 300 300 300 300 300 3	AND TO CLARK 1740 1740 1740 1870 GTGARAACCTAI CysLysFroll 1820 GTCCTGAACT ValLeuAsnL 1860 TCCASCAACT SerSerAsnL 1940 CTGATGGTTC LeuMetValH 2000 GCCAGSAGTS 2040 AGCTTGCACT	2010 2017 1930 1930 1930 1930 1950 1950 2010 2020 2020 2020 2020 2020 2020 20	TCACTGGGAGG UH1 SCys61u1 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAI eThrArg61ni 1900 GAGCATCTGG GAGCATCTGG USer I1eTrpi 1960 MACAGAAAAC 2020 SCTGGTAAGCA 2080 CCTTTAACCCT	2030 20290 20290 20290 20290 20290 20290 20290 20290 20191111111111	1900 ACTIC 1900 ACTIC In Phe 1860 CCTCC In Phe 1920 CCTCC 1920 CCTCC 1920 CCCC 1930 CCCCC 2040 ACCCC 2100 ACCCC
ACTATO ThrMet ATAGAT IleAsp GTGTCC ValSer TCABC SerAld GTTCAG	1750 39656A8CA6 39656A8CA6 39656A8CA6 39656A8CA6 3970 CA686CC66T 1870 CA686CC67 1870 CA686CC67 1870 CA686CG67 LysAlaAla 1930 1970 CTTCCT6CT6 2050 CAC56CTCC	ARTECOMPLE ARTSERASCI 1760 [1760 [1760 [1820 GTCCTGAACT ValLeuAsnL 1880 TCCAGCAACT SerSerAsnL 1940 CTGAGGAGTG 1940 GCCAGGAGTG 2000 GCCAGGAGTG 2060 AGCTTGCACT	101014001 1770 CCTGTTCTGG 1830 1830 1860GTTCCTa 1950 1560GTTCCT 1950 1950 1950 2010 GGGATGCAGG 2070 TACCCTTTTC	TCACTGGGAGA UH1 SCYSG1u' 1780 TACTAGATATI YThrArgTyr/ 1840 CACACGACAA eThrArgG1ni 1900 GAGCATCTGG USerIIETrpi 1960 MACAGAAAAC 2020 SCTGGTAAGCA 2080 SCTTTAACCCT	2030 2020 2020 2020 2020 2020 2020 2020	1940 (1900) CCTTC SINPhe 1980 CCTCCA 1980 TTCTC TTCTC 1980 2040 AGG64 2100 ACCG4
Phearc ACTATO ThrMet ATAGAT IleAs GTGTCC ValSer TCAGC SerAl GTTCA	1750 BAGTGAGCAG: SerGluGini 1810 (CAGACCCGTI GlnThrArgi 1870 CAGGCTGCT LysAlaAla: 1930 FACTTGACCI 1970 CTTCCTGCTG 2050 CACTGCTCC 2110	ARCTCCSARCT 1740 [S 1740 [S 1670AACCTAI CysLysFroTi 1820 GTCCTGAACT ValLeuAsnL 1880 TCCAGCAACT SerSerAsnL 1940 CTGATGGTTC LeuMetValH 2000 GCCAGGAGTG 2040 AGCTTGCACT 2120	101010400 1770 CCTGTTCTGG 1830 TGGGTCCCAT suG1YProll 1950 ATTGATGGAP 2010 GGGATGCAGE 2070 TACCCTTTTC 2130	TCACTGCGAGG UH1 SCys61u' 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAAA eThrArg61ni 1900 GAGCATCTGG USer I1eTrpi 1960 ICAGGAAAAAC 2020 ICTGGTAAGCA 2080 ICTTTAACCCT 2140	2030 2020 2020 2020 2020 2020 2020 2020	1900 ACTIC SIPhe 1800 ACTIC SIPhe 1980 CCTC4 LaSer 1980 CCTC4 LaSer 1980 CCTC4 LaSer 1980 CCTC4 LaSer 1980 CCC4 LaSer 2040 ACG64 ACC64 2100 ACC64 2120
ACTATO ACTATO ThrMet ATAGAT IleAsp GTGTCC ValSer TCAGC SerAld GTTCA	1750 1750 1967 19750 19750 1970 1070	ARTECOMPLE ARTSERASCI 1760 1760 1670 1820 GTCCTGAACT ValLeuAsnL: 1880 TCCAGCAACT SerSerAsnL 1940 CTGATGGTTC LeuMetValH 2000 GCCAGGAGTS 2060 AGCTTGCACT 2120 2060 2060 2120 2	101014001 1770 CCTGTTCTGG 1830 19808CCCAT 19809 19808CCCAT 1990 19808CTCCA 1950 ATTGATGGAG 2010 GGGATGCAGG 2010 GGGATGCAGG 2070 TACCCTTTC 2130	TCACTGGGAGG UH1 SCYSGIU' 1780 TACTAGATATI YThrArgTyr/ 1840 CACACGACACA eThrArgGIni 1900 GAGCACATCTGG USerIIeTrpi 1960 SCTGGTAAGCA 2080 CCTTTAACCCT 2140	2150 2070 2070 2070 2070 2070 2070 2070 20	1940 1800 CTTC ISAC CTTC ISAC
ACTATO ThrMet ATAGAI IleAsc GTGTCC ValSer TCAGC SerAl GTTCAG GTTCAG	1750 396T6A9CAG 396T6A9CAG 396T6A9CAG 396T6A9CAG 3970 5010 ThrArg 1970 5020 ThrArg 1970 5020 ThrArg 1970 5020 ThrAng 1970 5020 ThrAn	ARTECCSACT 1740 [S 1740 [S 16740ACCTAI CysLysFroTi 1820 6TCCTGAACT ValLeuAsnL 1860 TCCAGCAACT SerSerAsnL 1940 CTGATGGTTC LeuMetValH 2000 GCCAGGAGTS 2000 AGCTTGCACT 2120 TGCTGCTGAT	111 1170 1770 CCTETTCTGG 1830 TGGGTCCCAT 1950 TGGGTCCCAT 1950 TGGGTCCCAT 1950 TGGGTCCCAT 2010 GGGATGCAGE 2070 TACCCTTTTC 2130 TCTCAACTC	TCACTGCGAGE UNH SCYSGIU' 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAA eThrArgGIni 1900 GAGCATCTGG USERIIETrpi 1960 SCACGATCAGAAAAC 2020 SCTGGTAAGCA 2020 SCTGGTAAGCA 2080 SCTTTAACCCT 2140 380ACTTGTAA	2030 2030 2030 2030 2030 2030 2030 2030	1840 ACTTC 1860 CCTC 1860 CCTC 1860 CCTCC 1860 CCTCC 1920 CCTCC 1920 CCTCC 2040 ACGG 2100 ACCG 2160 CATG
Phear c ACTATE Thr Met ATAGA Il eAsc GTGCC Val Ser TCASC Ser Al A GTTCAS GTTCAS	1750 360 TBAGCAG 360 TBAGCAG 360 TBAGCAG 360 TAAGACCCGTI 361 Thr Ar g' 1870 360 TAAGACCCGTI 1970 370 TAACTTAACC 370 TATCTTCAA	ARTECOMPLE AnserAspL 1760 1760 1820 STCCTGAACTT ValLeuAspL 1880 TCCAGCAACT SerSerAspL 1940 CTGATGGAACT 2000 GCCAGGAGTG 2060 AGCTAGCACT 2120 TGCTGCTGAT	11111111111111111111111111111111111111	TCACTGCGAGG UH1 SCysGlu' 1780 TACTAGATATI yThr ArgTyr/ 1840 CACACGACAAA e Thr ArgGlni 1900 GAGCATCTGG GAGCATCTGGAAAAC 2020 SCTGGTAAGCA 2020 SCTGGTAAGCA 2080 SCTTTAACCCT 2140 389ACTTGTAA	2030 2070 2030 2030 2030 2030 2030 2030	1940 1800 CTTC ISAC CTTC ISAC CCTCA 1860 CCTCA 1920 TCCCA AGGGA 2100 ACCGA 2166 CATG
ACTATO ThrMet ATAGAI IleAsp GTGTCC ValSer TCASC SerAl GTTCAC GTTCA	1750 396T6A9CAG 396T6A9CAG 396T6A9CAG 396T6A9CAG 3970 501nThrArg 1970 5026T6CT -LysAlaAla 1930 74CTTT6ACC 2050 CACTGCTCC 2110 TATCTTTCAA	AND TO CLASH CONTRACT CONTRACT 1760 \$ 1760 \$ 1870 \$ 1820 \$ STCCTGAACT \$ 1820 \$ STCCTGAACT \$ 1880 \$ TCCAGCAACT \$ 1940 \$ CTGATGGTTC \$ 2000 \$ GCCAGGAGTS \$ 20040 \$ AGCTTGCACT \$ 2120 \$ TGCTGCTGAT \$	2010 20170 201770 201770 201770 201770 201770 201770 201770 201770 201770 2017 2010 2017 2010 20170 20	TCACTGCGAGE UNH SCYSGIU' 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAA eThrArgGIni 1900 GAGCATCTGG WSerIleTrpi 1960 MACCAGAAAAAC 2020 SCTGGTAAGCA 2020 SCTGGTAAGCA 2080 CCTTTAACCCT 2140 380ACTTGTAA	2030 2030 2030 2030 2030 2030 2030 2030	1840 ACTTC 1840 1840 CCTC 1840 CCTCC 1840 CCTCC 1840 CCTCC 1920 CCTCC 1920 CCTCC 2040 ACCCC 2040 ACCCC 2040 CACGG
Phearc ACTATE ThrMet ATAGAT 11 eAs GTGTCC Val Ser TCAGC Ser Al A GTTCAG BCTCAG	1750 BAGTGAGCAG SerGluGial 1810 CAGACCCGTI GinThrArg 1870 CAGGCCGCT LysAlaAla 1930 FACTTGACC ATHRLEUTHOL 1990 CTTCCTGCTG 2050 CACTGCCTCC 2110 TATCTTTCAA 2170	ARTSERASCIA ARTSERASCIA 1740 [S 1740 [S 1740 [S 1820 GTCCTGAACT ValLeuAsnL 1880 GTCCTGAACT SerSerAsnL 1940 CTGATGGTTC LeuMetValH 2000 GCCAGGAGTG 2040 AGCTTGCACT 2120 TGCTGCTGAT 2180	Initiation I	TCACTGGGAGG UH1 SCYSG1u ¹ 1780 TACTAGATATI YThr ArgTyri 1840 CACAGGACAAA eThr ArgG1ni 1900 GAGCATCTGG GAGCATCTGG USSer I1eTrpi 1960 CCTGGTAAGCA 2020 CCTGGTAAGCA 2020 CCTTTAACCCT 2140 300ACTTGTAA 2200	2030 2030 2030 2030 2030 2030 2030 2030	1800 ACTTC SnPhe 1860 CCTC4 1860 CCTC4 1980 TTCTC neLeu 1980 ACG6 ACG6 2100 ACCG4 2160 CCTG4 CCTG4
ACTATO ThrMet ATAGA1 IleAsp GTGTCC ValSer TCAGC SerAl GTTCAC GTTCAC	1750 396T6A3CAG 396T6A3CAG 396T6A3CAG 396T6A3CAG 3970 3970 3970 3970 3970 3970 3970 3970 3970 3070	ARTECLEMENT AnserAspL 1760 [S 16760 ASPL 16760 ASPL 1620 STCCTGAACT ValLeuAsnL 1680 TCCAGCAACT SerSerAsnL 1940 CTGATGGTTC LeuMetValM 2000 GCCAGGAGTG 2040 AGCTGCACT 2120 TGCTGCTGAT 2180 ASPL	UVAITYLE UVAITYLE I770 CTGTTTCTGG TGGGTCCCAT IS30 TGGGTCCCAT UGIYPF011 11 19 0 TGGGTCCAT 19 2010 GGGATGCAGG 2070 TCCCTTTTC 2130 TCTCAACTC 21700	TCACTGGGAG UHI SCYSGIU' 1780 TACTAGATATI YThr Ar g Tyri 1840 CACACGGACAA e Thr Ar gGIni 1900 GAGCATCTGGI USer II e Trpi 1940 MACAGAAAAAC 2020 CCTGGTAAGCA 2080 CCTTTAACCCT 2140 380ACTTGTAA 2200	2210 2210 2210 2017 2017 2017 2017 2017	1800 ACTIC SnPhe 1860 CCTCA 1850 CCTCA 1920 CCTCA 1920 CCCCA AGGGA 2100 ACCCA 2100 CATCA CATCA
ACTATE ThrMet ATAGAT 11 eAs GTGTCC Val Ser TCAGC SerAl GTTCAG GTTCAG	1750 300 TGAGCAG 300 TGAGCAG 300 TGAGACCCGTI 301 TCAGACCCGTI 301 TCAGACCCGTI 301 TCAGCCGCTGCT -LysAiaAla 1930 TACTTTGACC 3050 CACTGCCTCC 2110 TATCTTTCAA 2170 CTGTCTCTTT	1740 1 1740 1 1740 1 1670 1 1820 1 1820 1 1740 1 1820 1 1820 1 1820 1 1820 1 1820 1 1820 1 1820 1 1820 1 1940 1 2000 3 3000 3 2040 3 3000 3 2040 3 3000 3 2040 3 3000 3 2040 3 300 3 300 3 300 3 300 3 300 3 300 3 300 3 300 3 300 3 <tr td=""></tr>	111 111 11770 111 1130 111 1130 111 1130 111 1130 111 1050 111 1050 111 1050 111 1050 111 1050 115 2010 30647GCA86 2010 3070 TCACCCTTTTC 2130 1CTTCAAACTAC 2190 1CAAAATAA 100	TCACTGGGAGA UHI SCYSGIU' 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAA eThrArgGIni 1900 GAGCATCTGG GAGCATCTGG 2020 SCTGGTAAGCA 2020 SCTGGTAAGCA 2080 CCTTTAACCCT 2140 36AACTTGTAA 2200 TAAGTTTAAAAA	2017 - Laucional Control Contr	1800 ACTTC SnPhe 1860 CCTC4 1860 CCTC4 1980 TCCC4 AGG64 2100 ACCG4 2160 CATG6
Phearc ACTATE ThrMet ATAGA IleAs GTGCC ValSer TCAGC SerAld GTTCAG GTTCAG	1750 39656A8CA6 39656A8CA6 39656A8CA6 39656A8CA6 3975 3970 CAA86CCG6T 3970 CAA86CT6CT 4970 2050 2050 2050 2050 2050 2050 2050 2050 2050 2050 2110 TATCTTCAA 2170	ARTSERASDL 1760 1760 1760 1820 STCCTGAACT ValLeuAsnL 1880 TCCAGCAACT SerSerAsnL 1940 CTGATGGTTC LeuMetValH 2000 GCCAGGAGTS 2000 GCCAGGAGTG 2000 GCCAGGAGTG 2000 GCCAGGAGTG 2000 36CTTGCACT 2120 TGCTGCTGAT	101010400 1770 CCTGTTCTGG 1830 1830 1860GTCCAT 1950 TG6GGTTCCT 1950 ATTGATGGAG 2010 GGGATGCAGG 2070 TACCCTTTC 2130 TCTTCAACTG	TCACTGGGAGG UH1 SCYSGIU' 1780 TACTAGATATI YThr Arg Tyr/ 1840 CACACGACACA e Thr ArgGIni 1900 GAGCACATCTGGI USer II eTrpi 1960 SCTGGTAAGCA 2020 SCTGGTAAGCA 2080 CCTTTAACCCT 2140 389ACTTGTAA 2200	2210	18000 ACTTC SnPhe 18600 CCTCA 18507 1920 CCTCA 1980 TCCCA AGGGA 2100 ACCGG 2160 CATGG
ACTATE ThrMet ATAGAT 11 eAs GTGTCC Val Ser TCAGC Ser Al A GTTCAI GTTCAI GTTCAI	1750 300 TGAGCAG 300 TGAGCAG 300 TGAGACCCGTI 301 TCAGACCCGTI 301 TCAGACCCGTI 301 TCAGACCCGTI 302 TAAGGCTGCT 403 TACACTAGCC 403 TACACT	1740 1 1740 1 1740 1 1670 1 1670 1 1820 1 1820 1 1820 1 1820 1 1820 1 1820 1 1820 1 1820 1 1920 1 2000 3 2040 3 3 3 2040 3 3 3 2000 3 3 3 2040 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	111 111 11770 111 1130 111 1130 111 1130 111 1130 111 12950 111 2010 36686TTCCT 2010 36686TTCCC 2010 36686TGCAGC 2010 36686TGCAGC 2010 36686TGCAGC 2010 36686TGCAGC 2010 36687GCAGC 2010 36687GCAGC 2010 36687GCAGC 2010 36687GCAGC 2010 36687GCAGC 3000 3700 3000 3700 3000 3700 3000 3700 3000 3700 3000 3700 3000 3700	TCACTGGGAGG UNH SCYSGIU' 1780 TACTAGATATI yThrArgTyri 1840 CACACGACAAA eThrArgGIni 1900 GAGCATCTGG GAGCATCTGG 2020 SCTGGTAAGCA 2020 SCTGGTAAGCA 2080 CCTTTAACCCT 2140 3GAACTTGAAA 2200 TAAGTTTAAAAA	2017 - Laucional Control Contr	1800 ACTTC SnPhe 1860 CCTCA 1860 CCTCA 1980 TCCCA AGG64 2100 ACCG6 2160 CATG6

1150 1160 1170 1180 1190 1200 GGTGAACGAGACTGGATGTCCATAGTGACTCCTGCCAGGGATGGTCCCTGTGGGACAGTA GlyGluArgAspTrpMetSerIleValThrProAlaArgAsp6lyProCysGlyThrVal

FIG. 2. (B) Complete sequence of the insert of pRTHP1.

1988), we searched for such sites in genomic fragments that were just upstream of the rodent cDNA 5' end. Rat kidney nuclei were subjected to limited amounts of DNase I, then restricted for genomic Southern analysis. A genomic probe of 300 bp (derived by PCR) located in intron 2 was used (denoted in Fig. 3B as "P"). The expected 3 kb Hind III-Hind III fragment is reduced to 1.7 kb in a fraction of the nuclei (Fig. 4). This heterogeneity was anticipated because only about 10% of the cells in a whole kidney (those in the thick ascending limb and early distal nephron) are thought to express the THP gene. Importantly, the hypersensitive site is absent

from rat liver nuclei subjected to the same manipulations and also from DNA isolated from a rat kidney directly (without an intermediate nuclear fractionation and DNase I digestion step). We sequenced the locus of the hypersensitive site and found striking homology between the rat sequence and sequence obtained from a region of a human genomic clone just upstream of the alternatively spliced exon "a" reported by Pennica et al. A rat probe spanning the hypersensitive site (derived through PCR) was then used to locate the homologous region in the bovine genomic clones. This sequence comparison is presented in Fig. 5. The conserved



FIG. 3. Genomic organization of the bovine (A) and rat (B) THP gene. Boxes represent exons and have been numbered. The locus of the hypersensitive site in the rat clone has been marked as HS. Restriction sites are abbreviated: K, Kpnl; E, EcoRI; H, HindIII; X, Xbal; B, BamHI; N, Ncol.

1 2 3 4 5 6 M 7 8 9 10 11 12



FIG. 4. DNase I hypersensitive site analysis. Lanes 1 and 7: DNA from rat kidney and liver (respectively) isolated without an intermediate nuclear purification step so as to minimize degradation. Lanes 2–6: DNA from kidney nuclei subjected to DNase I at concentrations of 0, 1, 2, 5, and 10 μ g/ml. Lanes 8–12: DNA from liver nuclei subjected to 0, 1, 2, 5, and 10 μ g/ml of DNase I. All DNAs were digested with HindIII. The probe is from intron 1, a nonrepetitive, PCR-generated 300 bp fragment (see the Materials and Methods section). The original 3 kb HindIII-HindIII band (top arrow) is reduced to 1.7 kb (lower arrow) in approximately 10% of the sample and only in samples derived from kidney nuclei. Lane M is a marker lane, with a "1 kb ladder" containing 1 kb marker DNAs and a Hinf I digest of pBR322 (Gibco, BRL).

region spanned approximately 350 bp in all three species and contained distinctive promoter hallmarks: a TATA box and an inverted CCAAT box (underlined).

RNA Mapping to Locate the Transcription Start Sites

Primer extension analysis was performed on bovine outer medullary RNA. Two oligo probes, BTHP4 and BTHP7, randomly chosen from exon 2 produced primer-extension products at 16 nucleotides upstream from nucleotide position 1 of the cDNA containing plasmid pBTHP1 (Fig. 6A). An RNase protection assay using a 220 bp fragment (Ssp I-Pst I), which contained the first exon, gave a protected nucleotide size at a range of 32-33 confirming the first exon size of 31 nucleotides as shown by the primer extension assay (Fig. 7A). When primer extension analysis was performed on rat kidney RNA using two different antisense primers RTHPRE36 and RTH5, separated by 57 bases, major extension products of 102 and 159 nt were detected (Fig. 6B). RNase protection analysis using a 385 bp antisense riboprobe spanning the conserved region (from nt 229 to 614, Fig. 5) gave a protected fragment of 30 bp (Fig. 7B). Liver RNA was negative by both methods. The results of the mapping studies in both species were in close agreement and identified a cap site 32 bp and 34 bp downstream of the TATA boxes underlined in Fig. 5.

In Vitro Transcription

To date, and despite extensive searching, we have not found a cell line expressing THP message. To begin to study cell type specific regulation of THP, we elected to attempt in vitro transcription from a kidney extract, initially derived from whole rat kidneys. Similar extracts from other organs (e.g., liver) have proven to be valuable reagents for studying determinants of tissue-specific transcription in vitro. Our kidney extracts retain the capacity to utilize the in vivo cap site and are sensitive to low concentrations of α -amanitin (2.5 µg/ml). We found that the production of transcriptionally competent nuclear extracts from kidney required protease inhibitors in addition to those described by Schliber (Gorski et al., 1988), and the phosphatase inhibitor sodium molybdate (see the Materials and Methods section).

A rat THP promoter fragment from -1600 bp to -3 bp (relative to the assigned start site in the rodent gene, Fig. 5) was fused just upstream of the G-less cassette in pC₂AT (a kind gift of Dr. R. Roeder) to give pTHP380. Rat kidney nuclear extracts faithfully initiated transcription from the in vivo cap site (3 bp into the cassette) to produce the expected 380 bp transcript that is truncated by the chain-terminating guanosine nucleotide analog 3'-Omethyl GTP (lane 1, Fig. 8) In addition, the inclusion of α -amanitin at 2.5 µg/ml resulted in complete poisoning of transcription as is seen with other RNA Pol II promoters (lane 2, Fig. 8). A construct that has the orientation of the promoter fragment of pTHP380 reversed also gave no signal (data not shown). The results of this functional assay provide further proof that the 350 bp conserved region is the promoter for the THP gene.

DISCUSSION

One approach to the study of renal development is to define a hierarchical set of renal transcription factors that participate in a temporal and spatial cascade of transcriptional activation of kidney-specific genes, ultimately leading to the differentiated state. A tissue-specific marker gene can serve as a necessary springboard in such an approach. For example, much insight into the problem of hepatogenesis has been gained by studying the mechanisms that underlie liver-specific activation of the albumin gene (de Simone and Cortese, 1988).

As a first step towards elucidation of the mechanisms of transcriptional regulation of a kidney-specific gene, we have cloned and characterized the bovine and rat THP genes. Bovine and rat cDNAclones were obtained by screening kidney cDNA libraries with a human THP cDNA fragment. Genomic clones were then isolated by using the cDNAs as probes. The location of exons on the genomic map along with a detailed analysis of the exon/intron junctions was a necessary first step in delineating the boundaries of the THP transcription unit. The promoter was identified through a combination of strategies that included DNase I hypersensitive site analysis, interspecies sequence comparison, and RNA mapping studies. That the promoter fragment identified by these means directs RNA Pol II-mediated transcription in vitro further authenticates its central role in the transcription unit of the THP gene.

A cell line that expresses the THP mRNA at the normally high in vivo levels does not exist. Therefore, we have chosen to work with the bovine and rat genes to characterize gene regulatory elements by species comparisons, using transgenic mice and by biochemical approaches. The latter strategy requires large quantities of kidney nuclei that are readily available from animals. The cow kidney offers the additional advantage of further dissection of the

~ ~ ~								
16.15	213	273 276	340 331 335	868 395	155 150	199	69 64	25 25 20
				.,,	***	ທີ່ທີ່ທີ່	ഗഗഗ	999
< < U	* ৩ ৩ ৩	∢ ∪∪	+ ~ ~ ~	* < < <	.	e e 1)		
000	*	* 0 0 0			* 10 10 10			
ແ ບບ	* 0 0 0	* 0 0 0		* 0.0.0	* 5 5 5		* 0 0 0	≪00
			1.222			* 0 0 0	* < < <	* ৩ ৩ ৩
A 14 18	500			* 8 8 8	ن مر مر	* ~ ~ ~	нo	* ~ ~ ~
		*	****	* 0 0 0	* 8 8 8	* ৩ ৩ ৩	~ ~	000
HOH		* 0 0 0		* 6 6 6	00	* < < <	60	* ២២២
000	000	≪ ບ ບ	< ບບ	000	4 H A	000	€ €	* • • •
4 U F	« ••	U K K	⊢ ບ ບ	0 10 0	* 0 0 0	000	~ ~	* 0 0 0
* * * *	U	ບຂບ		000	* < < <	* < < <	۳ 00	* • • •
* 0 0 0	* • • •	* ৩ ৩ ৩	* 8 8 8	0 10 0	* < < <	* < < <	< H H	< U <
۲	⊢ ⊢	* ~ ~ ~	< ບ ບ	* 8 8 8	A U A	* ២២២	0 64	*
* 0 0 0	* 0 0 0	* 0 0 0	* 0 0 0	* • • •	< H U	* < < <	* < < <	* 0 0 0
U H U	0 + 0	* 0 0 0	* * * *	* • • •	* ৩ ৩ ৩	* 0 0 0	*	0 4 0
0 FP FP	* ២២២	* 0 0 0	* * * *	* < < <	* 0 0 0	A U A	E E ()	* 10 10 10
000	* = = =	* 0 0 0	< < 0	* 0 0 0	* * * *	£ 1) £		* ~ ~ ~
нно	UHU							
ບບ∢	URR		0.06					* ~ ~ ~
 < = ~ < < = ~ < < < < < < < < < < < < < < < < <	0 6 6							* 8 8 8
* 15 15 15	* ~ ~ ~			-			* ~ ~ ~	* ৩ ৩ ৩
		* ** ** **		+ 0 0 0	* 0 0 0	* * * *	* 6 6 6	* ৩ ৩ ৩
*		~ ~	* + + + +	* 6 6 6	< 00	H U U	* < < <	* ~ ~ ~

	~	< ⊷			~		~ •	¢υ	*	U	U U		~	υυ	*	~	~ ~	*	~ ~	< <		нυ	υ	~	~ (,	E				
	90	9 64			U		υu	20	*	н	••	*	U	υU	*	υ	υυ	*	••	• ••	*	ს ი	U	U	~ ~	: *	C	00		•	00
	н (່ວ		₽	н н		~ (טכ	*	υ	υυ	*	U	υυ	*	~	~ ~	*	00	0 0	*		ы н	• U	00	* (~		*	()	00
	υu	א כ			<		FH 6	J U		н	υυ	*	ы	н н	*	~	~ ~	*	• •			ک د	~	• ~	~ ~		E	0			~ ~
	E .	¢ 0		e	υυ		≪ E	i O		H	ю н	*	ы	н н	•	~	~ ~	*	00	່ວ	*	нн	ы ,	. 0	00	,				5	
	0	C F4		U	< ⊢		E C	ι		~	н н	*	υ	υυ			U	*		• ••		υυ	₽ ,		~ 4			0			
	н (טכ		U	υυ		00	່ວ		U	eυ		~	υυ		~	υυ		00	, o l		K H	~	0	0 6		ŭ		Ĵ	ž	2.2
	U U	9 64		ы	υH		~ (9 64		~	н н		υ	~ ~		H	0 0		UF			υυ	0	ō		,				5	
*	€4 6	H FH		υ	~ ~	*	~ ~	< <		υ			υ	αυ		~			ບັບ	, <u>.</u>	*	~ ~	Ā,	. e			2			2	0.0
	υ		*	U	ი ი	*	υυ	ט נ		ы		*	U	0 0		ы	нн		0 6				e .		-				-	-	
	H			~	υH		~			н	E+		~				(1) (1)		нн		•	at 13									0 44
	••	ט ד		~	ບັບ	*	00	սս	*	0	0.0	-	0	0.0		10	00	-				~ ~	2.1				0		*	H	H H
	υ	טנ	*	0	00		0.6	- 0		ň	- /h		ň		1	2		-										~ ~	-	0	00
	U F				ы (л					~	(n (n		2		1	2				171				. 0	00	. *	H	не		U	ە 🖍
	00	• · ·		0	() at			n / v		ž	20		~			2	~ ~		-			00	9	•	0 4		~	< 0	*	U	00
	н.		ير	~	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~				-	5		*	3	00					00	201		~ ~	«	~	0 4		H	·∺~.	*	4	~ ~
			-	2								*			*	A	~ ~	-#	нн	H	*	~ ~	< +	~	~ ~	• •	~	~ ~	*	~	~ ~
-	~			0			60	ی مەرد 		0	~ ~	*	~	~ ~		U U	U H	*	< <	· ~		~ ~	U	U	5 4	*	e	нн	*	۲	••
	2.		*	0	00		AC E	- 0		G	нн		~	0 >		υ	υн	*	~ ~	< <]	-	~ ~	≪ +	~	~ ~	: *	~	< <	*	U	ს ი
	RG (0	нн	*	00	90	*	~	~ ~	*	~	< <	*	U	ს ი	*	υυ	0	- H	9 9	ن 10	•	~ ~	*	e	н н	*	U	υυ
	U A			0	υн	*	e e			U				< <	*	H	H H	*	нн	• 🖬		< ບ	U	H	$\circ \circ$	*	~	< <	*	~	~ ~
	H (Эн	*	~	~ ~		υ	9 0	*	G	υυ	*	U	υυ		H	υυ	*	υv	0	- F	нн	₽ *	•	••	• *	H	нн	*	υ	υυ
*	υu	J U	*	G	თ თ	*	~ (90	*	<	~ ~	*	4	< <	*	S S	υυ	*	••	• 64		υн	E-I		U 🛪	1	c	~ ~`	*	H	нн
	e e		*	Ċ,	თ თ	*	~	<		G	υυ		~	< <		U I	H H		U <	: <		υн	υ			*	EH	H H	*	H	нн
		~		с С	нн		90	υυ		U	нн	*	υ	υυ	*	to I	••	*	~ ~	: <		ს ს	ن •	~	~ ~		~	< 0		₽	ს ი
*	с (9 (9	*	U I	υυ		С F		*	ы	H H	*	υ	υυ	*	U I	υυ	*	υυ	0		< ບ	≪ *	~	~ ~		U	e e	*	υ	ს ს
*	€ • E	- 6-	*	EI I	нн		€ •	C O	*	υ	υυ		~	⊢ <	*	e i	нн	*	υυ	0		чυ	υ	÷	υυ		~	ن که	*	U	υu
	« •	¢υ		~	ს ი		нι	> <		υ	ს ი		U	υυ		υ	υυ	*	υυ	ပါ	*	~ ~	< ∗	~	~ ~		υ	U H	*	Ē	н н н
	5 0	9 🗗		U I	••		E A	~		e	υυ		ы	0 6		U I	ны	*	υυ	0	* (υυ	υ *	~	~ ~		~	~ ~	*	U U	υυ
	H (0 0		~	υυ	*	υu	υυ		υ	ны		~	د ن	*	~	~ ~		υ	0	* 1		- + *	~	~ ~		U	U K		н н	ŏн
	< (0 ⊢		U.	υн		U			H	υυ		~	ບຂ	*	~	~ ~			Ū.	* 1	ບບ	υ 		0 4					н	
*	~ ~	< <	*	E I	н н		€1 E	ιu		н	υυ	*	υ	υυ	*	0	0 0	*	~ ~		* (ບັບ	- U	н	at 19		ē	00			0.6
	υι	0 0		U I	нυ	*	€+ 6			ы	ს ს		H	υн		~	00	*	00			at ()	<u> </u>	н	0.0			00		~	0.0
	U F			U I	н 0	*	υι	υu		ы	~ ~		ы	00			сн.			Ы			- -		22					<i></i>	
	н е	-		£4 -	υH		≪ €		*	υ	υυ		H				<u>م</u> ا		0 0			0.0	0 4	0	00		ž	0.0		2	
	00	, 0		U I	ны		Æ	- υ		Ū,	ບ່ອ	÷.	H	нн		E I	U FL		ны	Ц,					00		5	~		2	
	0.4	C	*	0	0 0			0		ŏ,	о н			0.0		~	0.01		11 10			22	2				-	~ 0			
	ΕL	່ວ່	*	н	ны			ō		ō		-	ē.	00		2			a 10				~ -		20		2			2	00
*	0.0	5.0		E I	0.0		0.0	n (n		ě.	-		ŭ	20		2			~ ~												
	E E	- 64	*	-									10	10 10		2		Ē	2.0			22	2 3	-						5	
	a 1			0	· · ·								ž	22						. <u> </u>					-	1 *	0	6.6		R.	00
	2		-			-	~ 1					*	2			2	22						-	0	00			AG, AG,		A.	
				2			~ ~						5										.			1 1	0	00		H	00
	-						00			-			5			5					* *		AC #	0	00		~	00		H	0 43
		. ~.										*	2			H		*	00			3 H	÷ ن	E4	HH	*	G	00	*	~	~ ~
	00	-					- C - R					*		R., R.,		- A			0 44			0	. *	•	нн			0 6		G	0 0
					RG, RG,			G 14G				*	0	00		~			HO	0			« *	~	< <			0 14		0	υυ
	00		#			*	00	0		ن ن ح		*	0		*	HI	1	*	00	0	* *		•	H	< H	1 *	~	~ ~		4	H 4
	~ ~		#	HI	HH			~		~ 1	нн		н	е 0	*	HI	ны	*	нн	E E	* •	< <	≪ *	~	< <	1	~	00		~	< U
	00	0		0	00			0	*	0	00		H	₽	*	0	9 0		U H	E E	* *	< <	≪ *	~	~ ~	•	~	~ ~	*	ž	~ ~
	HE	• 0	*	ы	ны			H		« 1	HO	*	н	нн	*	~	۱~ ۲		< H	H	* *	< <	≪ *	~	~ ~	1 *	G	00	(છ	00
	θC	0 0	*	ы	нн	*	н e			U I	нн		o	~ ~	*	0	0 0	*	~ ~	<	* (90	U		ک م	1		~		H	E U
	« F	• •		~ (⊢∢		~ (5	H 4			H	*	0	υu	*	e e	E E	* (90	v *	~	< <	*	U	00)ن	بر و
*	9 9	0 0	*	0	υυ		U C	0 0		6	U			U		0	່ວ	*	υu	U U		~	≪ *	×	~ ~			00		e i	Ėυ
	€ €	• •		G	υ		O F	- 6-		U			U	U U		< 1	H H	*	н н	÷	(9 0	+ +	U	თ თ	*	<	~ ~	*	e I	н н
	••	4		C	U			U	*	e (H H		H	нн	*	ы	H H	*	нн	•		< <	E+ *	G	υυ	•	υ	0 0		υ	υĸ
	υυ	,	*	~	< <			U		9	< <		U	нн	*	H (H H	*	υυ	υ	6	υ		U	9 <	*	U	0 0		~	αU
	€+ E	4		U,	< ບ		υu	0 0	*	<	< <		EH -	чU		e (ບບ	*	υυ	υ	6	υ		~	ы 0	*	H	н н	*	ы	нн
	e e	0 0	*	ы	-		۲ ک	- 6-		U i	0 ک	*	υ	υυ		н (່ວບ		ບ <	~	* *	~ ~	~	~	ن 🔉	*	H	H H		υı	⊢ ≪
		< ◄		H		*	~ ~	< ◄		0	U FH		U.	υH	*	н	ны	*	ს ს	U	.* E	-	H +	~	~ ~	*	~	< <	*	υ	υυ
	≪ F	4		υ			0 4	< <		υ (00		Ċ	υυ		н (ບບ		~ ~	H	* (υυ	υ	~	~		υ	~ <	*	<	~ ~
	< (0 0		υ				U			< <			υo	*	~	ا~ >	-	υu	U	c	> <	ب د	υ	υυ		н	нн	*	υi	υυ
	<u>م</u> م	. 0.		ο, ο	<u>ъ</u> , <u>ъ</u> ,		م م	. 0.		A 1	<u>ъ</u> , <u>ъ</u> ,		ρ.	ል ል		А (а, с ,		44	.	6		אס	Δ,	م م		Δ,	8		<u>.</u>	<u>a</u> a
	E	H		H			TH	E		Ë			E	H H		E			ΞE	TH I		3 8	HI	H	표표		H	33		胃	33
	ים	1 4		ם	ਸੰਸ		ומ	4 Æ		ם	нд		A	нż		<u>م</u>	ня		дн	F	Ĩ	ч.	c	۵	нд		۵	чд		ַם	нд

5050

116 108 107

FACING PAGE

FIG. 5. Sequence alignment of promoters from three species: rat, cow, and human. The start sites are circled. TATA and inverted CCAAT boxes have been underlined, as have conserved regions at nucleotides 258-300, 359-400, and 468-487.

outer medulla to enrich for tissue-containing THP message at high amounts, and we have recently also successfully produced in vitro transcription competent extracts from this source (data not shown). The kidney nuclei served a twofold purpose. First, the nuclei are a source of the THP gene in its natural chromatin context; this allowed us to identify a kidney-specific DNase I hypersensitive site and to thereby locate the promoter. Second, the nuclei are a source of nuclear protein that could be used to carry out DNA binding studies (data not shown) and in vitro transcription.

The identification of the cap site proved complicated. The rat cDNA clone rTHP-1 ends to 16 bp upstream of the initiator methionine codon. Primer extension with two gene-specific primers indicated that the rat cap site was 68 bp upstream of the initiator codon. The rat extension product was not likely to be contiguous with the genomic sequence due to two reasons: (1) RNase protection analysis using an antisense riboprobe in this region gave a protected fragment that had its 5' end only 39 bp upstream of the initiator codon (data not shown); this region has a consensus splice acceptor sequence; (2) the sequences upstream of the putative cap site-were the primer extension product to be contiguous with the genomic sequence-might have contained easily identifiable promoter elements. Because this was not the case, it seemed likely that a small exon having a maximum length of 30 bp existed somewhere upstream in the gene.

Pennica et al. (1987) isolated two classes of human THP cDNAs. The first class has a 25 bp exon that they refer to as exon "a" at the 5' end. The second class has as its 5' end the region just upstream of and contiguous with the second exon of the first class of cDNA. They refer to this as exon "b" and propose an alternative splicing event that gives rise to the two classes. We performed primer extension analysis using a human exon 2-specific primer on human RNA and found two extension products that agree with the boundaries of the cDNA ends of Pennica et al. (data not shown). However, both the rat and bovine THP mRNAs are associated with only one 5' end. We think that this speciesspecific variation arises because a fraction of the human THP RNAs retain a region as the 5' end (due to an abnormal splicing event) that in the other two species is an intron. We favor this explanation over the possibility that the human gene is associated with an additional promoter located upstream of exon "b" because there are no well-conserved promoter elements there and because that region does not serve as a promoter in vitro (data not shown).

For these reasons, it seemed likely that the expected 30 bp first exon in the rat gene would be the homologue of the human exon "a." If this were indeed the case, we expected extensive interspecies homology in the putative promoter regions upstream of these first exons. However, it proved difficult to locate the homologue of exon "a" on the rat gene by hybridization, due to the small size of the oligomer and its strong hairpin structure. Instead, the clue to the location of the rat homologue of exon "a" came from the mapping of a kidney-specific DNase I hypersensitive site, located approximately 700 bp upstream of the splice acceptor site of the rat exon 2. Importantly, DNA sequencing of a 350 bp region covering the hypersensitive site revealed an 80% homology with the genomic sequence upstream of human exon "a." The bovine gene contains the same conserved 350 bp region, and the homology between the bovine and the rat region is 72% (see Fig. 5).

RNase protection analysis identified a 5' end in the rat gene that agreed precisely with the primer extension result and that accounted for the 30 additional bases (see Fig. 7). The bovine primer extension and RNase protection results further agreed with the rat RNA mapping.

The promoter region is associated with a TATA element at -34 and an inverted CCAAT element at -65. In addition, three highly conserved elements were seen at rodent nucleotides 258–300, 359–400, and 468–487 (underlined in Fig. 5).

We wished to see whether the THP promoter would function in an in vitro transcription system because the absence of a cell line prevented us from transient assay testing. A G-less cassette reporter (pTHP 380) gave an accurate size transcript of 380 bp. Furthermore, the inclusion of α -amanitin at 2.5 µg/ml completely abolished transcription, as would be expected of a Pol II promoter. These results show that the conserved region can serve faithfully as a Pol II promoter in an in vitro system. Future experiments will examine the role of the individual conserved elements through the use of this in vitro system. Ultimately, reporter constructs in transgenic mice will reveal the precise elements required for high-level, kidney-specific expression of the THP





FACING PAGE

FIG. 6. (A) Primer extension of bovine kidney outer medullar RNA ($20 \ \mu g$) (lane 1) and tRNA (lane 2) with primers, bTHP4 and bTHP7 (not shown), randomly chosen from exon 2. The major extension product identified is 16 nucleotides upstream from the 5' end of bTHP1. A sequencing ladder was produced with bTHP1 and the same oligo primers. (B) Primer extension analysis of rat kidney mRNA ($2 \ \mu g$) with primers RTHPRE36 (lane 1) and RTH5 (lane 2). Lanes 3 and 4 are extensions using rat liver mRNA ($2 \ \mu g$) and primers RTHPRE36 and RTH5, respectively. The major extension products (102 nt in lane 1 and 159 nt in lane 2) are marked by arrows. A sequencing ladder was run next to the reactions for sizing purposes.

message. With the cloning of THP genomic fragments containing all the exons as well as the promoter and considerable 5' flanking sequence, we anticipate performing these experiments in the future. Finally, because virtually nothing is known about kidney-specific gene expression and nephron segment-specific expression, the studies reported here initiate a data base on this topic with implications to understanding kidney growth and differentiation.

ACKNOWLEDGEMENT

This work was supported by NIH DK44921 to V.P.S.

REFERENCES

- R. Breatnach and P. Chambon (1981), Annu Rev Biochem 50, 349.
- P. Chomczynski and N. Sacchi (1987), Anal Biochem 162, 156-159.
- Current Protocols in Molecular Biology (F. Ausubel, ed.) (1991), Green Publishing Associates and Wiley-Interscience.
- V. DeSimone and R. Cortese (1988), Oxford Surveys on Eukaryotic Genes 5, 51–90.
- A. P. Feinberg and B. Vogelstein (1983), Anal Biochem 132, 6-13.
- A. P. Fletcher (1972), in Glycoproteins: Their Composition, Structure and Function (A. Gottschalk, ed.), Elsevier, New York, pp. 892–908.
- K. Gorski, M. Carneiro, and U. Schliber (1986), Cell 47, 767–776.
- D. S. Gross and W. T. Garrard (1988), Annu Rev Biochem 57, 159–197.

- C. Hession, J. M. Decker, A. P. Sherblom, S. Kumar, C. C. Yue, R. J. Mattaliau, R. Tizard, E. Kawastima, S. Scheissmer, S. Heletky, E. P. Chow, C. A. Burne, A. Shaw, and A. V. Muchmore (1987), Science 237, 1479–1484.
- A. V. Muchmore and J. M. Decker (1985), Science 229, 479.
- D. Pennica, W. J. Kohr, W. Kuang, D. Glaister, B. B. Aggarwal, E. Y. Chen, and D. V. Goedder (1987), Science 236, 83–88.
- J. Sambrook, E. F. Fritsch, T. Maniatis (1989), Molecular Cloning, A Laboratory Manual, 2nd ed. (N. Ford, ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- E. A. Schenk, P. H. Schwartz, R. A. Lewis (1971), Lab Invest 25, 92.
- I. Tamm and F. L. Horsfall (1950), Proc Soc Exp Biol Med 74, 108.



B M 1 2 3



FIG. 7. (A) RNase protection assay for the bovine THP. Lane 1: bovine outer medulla RNA. Lane 2: tRNA. EcoRI-Pst I 800 bp genomic fragment that contains the first exon was subcloned into pBluescript II. The plasmid was digested with Ssp I, and the 220 bp antisense fragment was transcribed and used for hybridization. The arrow indicates a DNA nucleotide size range of 32–33, consistent with the first exon size of 31 nt, as shown by the primer extension assay. Bars indicate nucleotide sizes of 711, 489, 404, 364, 242, 190, 147, 118, 110, 67, 57, 34, 26, from the top. (B) Rnase protection analysis of rat samples using an antisense riboprobe from -350 to +35. Lane 1: liver total RNA (10 μ g). Lane 2: kidney total RNA (10 μ g). Lane 3: 10 μ g tRNA. The top arrow in the marker lane points to 34 nt; the bottom arrow to 26 nt. The major transcript end (arrow in lane 2) is at 30 nt.

M 1 2

FIG. 8. In vitro transcription using kidney nuclear extracts. A total of 1 μ g of template and 60 μ g of nuclear protein were used for both reactions. Lane 1: pTHP380. Lane 2: pTHP380 with 2.5 μ g/ml α -amanitin.

